**📑 Informe de Sobremuestreo**

**Configuración:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Dataset *Glass*** | | | | | |
| Densidad | Riesgo | Pureza | Factor de equilibrio | Random state | Modo espacial |
| 75 | 50 | Entropía | 0.8 | 42 | 2D |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **🔎 Clase 1** | | | | | | | | |
| Instancias originales | Objetivo balance | Estado | Motivo | Muestras sintéticas | | Candidatas | Filtradas | Riesgo medio/ densidad/ entropía/ vecinos validos |
| 56 | 48 | No se sobremuestrea | Ya superaba el objetivo (56 ≥48) | 0 | | **-** | **-** | No aplica |
| **Umbrales aplicados** | | | | | | | | |
| Densidad | | | | | Entropía | | | |
| **-** | | | | | **-** | | | |
| **Ratios** | | | | | | | | |
| Original | | | | | Resampled | | | |
| 0.327 | | | | | 0.327  👉 Esta clase ya estaba suficientemente representada, por lo cual no se generaron sintéticas. | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **🔎 Clase 2** | | | | | | | | |
| Instancias originales | Objetivo balance | Estado | Motivo | Muestras sintéticas | | Candidatas | Filtradas | Riesgo medio/ densidad/ entropía/ vecinos validos |
| 61 | 48 | No se sobremuestrea | Ya superaba el objetivo (61 ≥48) | 0 | | **-** | **-** | No aplica |
| **Umbrales aplicados** | | | | | | | | |
| Densidad | | | | | Entropía | | | |
| **-** | | | | | **-** | | | |
| **Ratios** | | | | | | | | |
| Original | | | | | Resampled | | | |
| 0.327 | | | | | 0.327  👉 Esta clase ya estaba suficientemente representada, por lo cual no se generaron sintéticas. | | | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **🔎 Clase 3** | | | | | | | | | | |
| Instancias originales | Objetivo balance | Estado | Motivo | Muestras sintéticas generadas | | Candidatas | Filtradas | Riesgo medio/ densidad/ entropía/ vecinos validos | Densidad media | Vecinos validos promedio |
| 14 | 48 | Sobremuetreada | candidatas válidas presentes | 34 | | Suficientes | Algunas eliminadas | 0.72 (moderado, aceptable) | Alta, vecinos estables | Adecuados para interpolación |
| **Umbrales aplicados** | | | | | | | | | | |
| Densidad | | | | | Entropía | | | | | |
| 1.0 | | | | | 0.72 (por debajo del corte critico) | | | | | |
| **Ratios** | | | | | | | | | | |
| Original | | | | | Resampled | | | | | |
| 0.082 | | | | | Aumentado (total dataset 205) 👉 Fue la **única clase minoritaria que PC-SMOTE reforzó con éxito**, gracias a su densidad y riesgo moderados. | | | | | |

**🔎 Clase 5**

* **Instancias originales:** 10
* **Objetivo balance:** 48
* **Estado:** *No se sobremuestrea*
* **Motivo:** *insuficientes\_filtradas* → todos los candidatos descartados.
* **Muestras sintéticas generadas:** 0
* **Riesgo medio:** 0.91 (alto)
* **Densidad media:** baja
* **Entropía:** 0.91 (mezcla alta de vecinos de otras clases)
* **Vecinos válidos promedio:** muy pocos  
  👉 Esta clase se encontraba en **zonas ruidosas**, con ejemplos mezclados con otras clases y escasa densidad de minoritarios → descartada.

**🔎 Clase 6**

* **Instancias originales:** 7
* **Objetivo balance:** 48
* **Estado:** *No se sobremuestrea*
* **Motivo:** *insuficientes\_filtradas* → todos los candidatos descartados.
* **Muestras sintéticas generadas:** 0
* **Riesgo medio:** 0.97 (muy alto, casi ruido)
* **Densidad media:** muy baja
* **Entropía:** 0.97 (extrema mezcla con otras clases)
* **Vecinos válidos promedio:** prácticamente nulos  
  👉 Caso más crítico: instancias aisladas, dispersas y dominadas por vecinos de otras clases. PC-SMOTE las consideró **ruido / outliers** y evitó generar sintéticas.

**⚙️ Parámetros globales del experimento**

* **Técnica:** PC-SMOTE
* **Factor de equilibrio:** 0.8 (cada minoritaria buscaba alcanzar el 80% del tamaño de la mayoritaria)
* **Random State:** 42 (reproducibilidad)
* **Modo espacial:** 2D (proyección vectorial del dataset)
* **Total dataset:**
  + Original: 171 instancias
  + Resampled: 205 instancias

**📌 Conclusión**

* **Clases 1 y 2:** no se tocaron porque ya superaban el objetivo.
* **Clase 3:** única beneficiada con 34 sintéticas → pasó de 14 a 48 instancias.
* **Clases 5 y 6:** descartadas por **alto riesgo, alta entropía y baja densidad**, consideradas zonas ruidosas.
* El sobremuestreo fue **selectivo**, priorizando seguridad y evitando introducir ruido artificial.